# **PROJECT: Build a detector for Phishing websites (LR)**

The purpose of this document is to specify the requirements for the project “Build a detector for Phishing websites (LR)” Apart from specifying the functional and non-functional requirements for the project, it also serves as an input for project scoping.

**Project Objective**

The purpose of the project is to use one or more of the classification algorithms to train a model on the Phishing website dataset in order to train a model which can detect phishing websites.

**Project Description and Scope**

You are provided with the following resources that can be used as inputs for your model:

1. A collection of website URLs for 11000+ websites. Each sample has 30 website parameters and then a class label identifying it as a phishing website or not (1 or -1).

2. Code template containing these code blocks :

a. Import modules (part 1)

b. Load data function + input/output field descriptions

You are expected to write the code for a binary classification model (phishing website or not) using Python Scikit-Learn that trains on the data and calculates the accuracy score on the test data.

**Project Guidelines**

Begin by extracting the ipynb file.

Exercise 1 : Build a phishing website classifier using Logistic Regression with “C” parameter = 100.

Use 70% of data as training data and remaining 30% as test data.

(hint : use Scikit-Learn library LogisticRegression)

(hint : look at the Logistic Regression tutorial taught earlier in the course)

Print count of misclassified samples in the test data prediction as well as the accuracy score of the model.

Exercise 2 : Train with only two input parameters. e.g. parameter index 5 (parameter Prefix\_Suffix) and 13 (URL\_of\_Anchor) only out of 30 parameters. Check accuracy on the test data again and compare the accuracy with the previous value, where it was trained on all the parameters.

Plot the test samples along with the decision boundary when trained with index 5 and index 13 parameters.

**Pre-requisites**

To execute this project you will need:

* Python 3.5
* Scikit-Learn library
* Jupyter notebook

**Programming hint for the student :-**

For ideas, look at tutorials accompanying various modules taught in the course.